(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/34803 A2

(51) International Patent Classification⁷: C12N 15/12, C07K 14/47, 16/18, C12Q 1/68, G01N 33/574, A61K 67/027

(21) International Application Number: PCT/US00/30945

(22) International Filing Date:
9 November 2000 (09.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/438,564 10 November 1999 (10.11.1999) US

(71) Applicant: ZYMOGENETICS, INC. [US/US]; 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(72) Inventors: HOLLOWAY, James, L.; 833 N.E. 89th Street, Seattle, WA 98115 (US). GAO, Zeren; 9502 179th Pl. N.E. #3, Redmond, WA 98052 (US). WHITMORE, Theodore, E.; 6916 152nd Avenue N.E., Redmond, WA 98052 (US).

(74) Agent: JOHNSON, Jennifer, K.; ZymoGenetics, Inc., 1201 Eastlake Avenue East, Seattle, WA 98102 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regionally): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL CRIB PROTEIN ZMSE1

(57) Abstract: The present invention relates to polynucleotide and polypeptide molecules for zmse1, a novel human CRIB protein. The polypeptides, and polynucleotides encoding them, may be used for detecting human chromosomal abnormalities and cancers. The present invention also includes antibodies to the zmse1 polypeptides.

WO 01/34803 A2

Description

NOVEL CRIB PROTEIN ZMSE1

BACKGROUND OF THE INVENTION

The Ras family of proteins is comprised of small GTPases that are subdivided into several sub-families known to be involved in diverse cellular actions, such as cell proliferation, differentiation and apoptosis. Moreover, Ras is a known oncogene, and the Ras protein is implicated in oncogenic cell transformation through complex signaling pathways that employ an increasing number of downstream effectors. Some of these effectors are included in Ras sub-families, such as, for example, the Rho subfamily of small GTPases.

The Rho family proteins are implicated in regulating diverse cellular processes as well. One prominent Rho activity comprises effecting actin cytoskeletal organization. Such activities include regulating cell shape, cell attachment and adhesion, cell motility and invasion, cell-cell interactions, cell proliferation, differentiation and apoptosis. For example, Rac1, RhoA and Cdc42 are all implicated in promoting cell motility and invasion. As such, these proteins may be involved in promoting motility, invasiveness and metastasis of tumor cells. In addition, disassembly of actin stress fibers is associated with malignant transformation. Moreover, another prominent and distinct Rho activity includes activation of signaling cascades that enhance gene expression through the induction of various transcription factors, resulting in cell proliferation, cell cycle progression, differentiation and apoptosis. For Example, the Rho proteins Rac1 and Cdc42 activate Jun NH_2-terminal kinases (JNKs), which in turn activate Jun, ATF-2 and Elk-1 nuclear transcription factors. Rho family proteins can also activate NFκB and SRF transcription factors. Virally transduced and mutated versions of cellular Fos, Jun, and NFκB were originally identified as potent retroviral oncogenes implicated in various tumors and cancerous states (Bishop, J.M., Cell 64:235-238, 1991). Thus, Rho family mediated changes in

gene expression likely contribute to their proliferative actions, and play a role in cell transformation and cancer. Moreover, several Rho family proteins have been shown to be important for Ras transforming activity. For reference, see Zohn, I.M. et al., Oncogene 17:1415-1438, 1998; Maruta, H. et al., Microsc. Res. Tech. 47:61-66, 1999; Aspenstrom, P. Exper. Cell. Res. 246:20-25, 1999; Banyard, J. and Zetter, B.R., Cancer and Metast. Rev. 17:449-458, 1999; and, Michiels, F. and Collard, J.G., Biochem Soc. Symp. 65:125-146, 1999.

In addition, several effector proteins are known to bind Rho family members, such as Cdc42 and Rac. The effectors that bind Cdc42/Rac have a consensus binding motif designated the Cdc42/Rac Interactive Binding (CRIB) motif (Burbelo, P.D. et al., J. Biol. Chem. 270:29071-29074, 1995). These effectors show GTP-dependent interaction with Cdc42 and/or Rac1, and may or may not show kinase activity. For Example, the Cdc42 effector, MSE55 is a non-kinase effector that specifically binds Cdc42 in a GTP-dependent manner, is localized to membrane ruffles, and induces long actin-based protrusions or cellular extensions in fibroblast cells (Burbelo, P.D. et al., Proc. Natl. Acad. Sci. USA 96:9083-9088, 1999). For other references on Cdc2/Rac and their effects on membrane ruffling, actin stress fibers, lamellipodia and the like, see, for example, Ridley, A.J. et al., Cell 70:401-410, 1992; and Nobes, C.D., and Hall, A. Cell 81:53-62, 1995. Other CRIB proteins are implicated in human disease, such as the Wiskott-Aldrich Syndrome, which is an X-linked recessive disorder characterized by thrombocytopenia, recurrent infections due to defective T- and B-cell function, and eczema. The CRIB protein Wiskott-Aldrich Syndrome Protein (WASP) is mutated in this disease, and it is also a Cdc42 effector (Symons, M. et al., Cell 84:723-734, 1996). Because these CRIB proteins influence members of the Rho family of proteins, these effectors may also play a role in cell proliferation, transformation, motility and metastasis. For reference, see Zohn, I.M. et al., supra.

Considering the importance of this family of proteins, there is a continuing need to discover new Ras and Rho family members and their effector proteins that modulate the cytoskeleton, actin polymerization, cell motility and invasion, and the like, and affect proliferation, differentiation, transformation,

metastasis and apoptotic pathways. The *in vivo* activities of both inducers and inhibitors of these pathways illustrate the enormous clinical potential of, and need for, such novel proteins, their agonists and antagonists, for example, in cancer therapy. The present invention addresses this need by providing such polypeptides for these and other uses that should be apparent to those skilled in the art from the teachings herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a hydrophobicity plot of zmse1 using a Hopp/Woods hydrophilicity profile based on a sliding six-residue window, with buried G, S, and T residues and exposed H, Y, and W residues ignored.

Figure 2 is an alignment of human zmse1 (zmse1) (SEQ ID NO:2), and mouse zmse1 (MUZMSE) (SEQ ID NO:5).

DESCRIPTION OF THE INVENTION

In one aspect, the present invention provides an isolated polynucleotide that encodes a zmse1 polypeptide comprising a sequence of amino acid residues that is at least 90% identical to an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 147 (Ala); (b) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 336 (Ser); (c) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 356 (Ser); and (d) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val), wherein the amino acid percent identity is determined using a FASTA program with ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62, with other parameters set as default. In one embodiment, the isolated polynucleotide disclosed above is selected from the group consisting of: (a) a polynucleotide sequence as shown in SEQ ID NO:1 from nucleotide 199 to nucleotide 639; (b) a polynucleotide sequence as shown in SEQ ID NO:1 from nucleotide 640 to nucleotide 1206; (c) a polynucleotide sequence as shown in SEQ ID NO:1 from nucleotide 640 to nucleotide 1266; and (d) a polynucleotide sequence as shown in SEQ ID NO:1 from nucleotide

199 to nucleotide 1266. In another embodiment, the isolated polynucleotide disclosed above comprises nucleotide 1 to nucleotide 1068 of SEQ ID NO:3. In another embodiment, the isolated polynucleotide disclosed above encodes a polypeptide that comprises a sequence of amino acid residues selected from the group consisting of: (a) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 147 (Ala); (b) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 336 (Ser); (c) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 356 (Ser); and (d) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val).

In a second aspect, the present invention provides an expression vector comprising the following operably linked elements: a transcription promoter; a DNA segment encoding a zmse1 polypeptide as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val); and a transcription terminator, wherein the promoter is operably linked to the DNA segment, and the DNA segment is operably linked to the transcription terminator.

In a third aspect, the present invention provides an expression vector as disclosed above, further comprising a secretory signal sequence operably linked to the DNA segment.

In a fourth aspect, the present invention provides a cultured cell comprising an expression vector as disclosed above, wherein the cell expresses a polypeptide encoded by the DNA segment.

In another aspect, the present invention provides a DNA construct encoding a fusion protein, the DNA construct comprising: a first DNA segment encoding a polypeptide comprising a sequence of amino acid residues selected from the group consisting of: (a) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 147 (Ala); (b) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 336 (Ser); (c) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 356 (Ser); (d) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 337 (Arg), to amino acid number

356 (Ser); and (e) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val); and at least one other DNA segment encoding an additional polypeptide, wherein the first and other DNA segments are connected in-frame; and wherein the first and other DNA segments encode the fusion protein.

In another aspect, the present invention provides an expression vector comprising the following operably linked elements: a transcription promoter; a DNA construct encoding a fusion protein as disclosed above; and a transcription terminator, wherein the promoter is operably linked to the DNA construct, and the DNA construct is operably linked to the transcription terminator.

In another aspect, the present invention provides a cultured cell comprising an expression vector as disclosed above, wherein the cell expresses a polypeptide encoded by the DNA construct.

In another aspect, the present invention provides a method of producing a fusion protein comprising: culturing a cell as disclosed above; and isolating the polypeptide produced by the cell.

In another aspect, the present invention provides an isolated polypeptide comprising a sequence of amino acid residues that is at least 90% identical to an amino acid sequence selected from the group consisting of: (a) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 147 (Ala); (b) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 336 (Ser); (c) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 356 (Ser); and (d) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val), wherein the amino acid percent identity is determined using a FASTA program with ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62, with other parameters set as default. In one embodiment, the isolated polypeptide disclosed above comprises a sequence of amino acid residues selected from the group consisting of: (a) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 147 (Ala); (b) the amino acid sequence as shown in SEQ ID NO:2 from

amino acid number 148 (Asn), to amino acid number 336 (Ser); (c) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 148 (Asn), to amino acid number 356 (Ser); and (d) the amino acid sequence as shown in SEQ ID NO:2 from amino acid number 1 (Met), to amino acid number 356 (Val).

In another aspect, the present invention provides a method of producing a zmse1 polypeptide comprising: culturing a cell as disclosed above; and isolating the zmse1 polypeptide produced by the cell.

In another aspect, the present invention provides a method of producing an antibody to zmse1 polypeptide comprising: inoculating an animal with a polypeptide selected from the group consisting of: (a) a polypeptide consisting of 13 to 343 amino acids, wherein the polypeptide is identical to a contiguous sequence of amino acids in SEQ ID NO:2 from amino acid number 1 (Met) to amino acid number 356 (Val); (b) a polypeptide as disclosed above; (c) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 337 (Arg) to amino acid number 356 (Val); (d) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 96 (Glu) to amino acid number 101 (Asp); (e) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 226 (Asp) to amino acid number 231 (Asp); (f) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 346 (Met) to amino acid number 351 (Glu); (g) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 360 (Arg) to amino acid number 365 (Glu); (h) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 347 (Asp) to amino acid number 352 (Asp); and (i) a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 from amino acid number 348 (Glu) to amino acid number 353 (Glu); and wherein the polypeptide elicits an immune response in the animal to produce the antibody; and isolating the antibody from the animal.

In another aspect, the present invention provides an antibody produced by the method disclosed above, which binds to a zmse1 polypeptide. In one embodiment, the antibody disclosed above is a monoclonal antibody. In another aspect, the present invention provides an antibody which specifically binds to a polypeptide as disclosed above.

In another aspect, the present invention provides An antibody which specifically binds to a polypeptide disclosed above.

In another aspect, the present invention provides a method of detecting, in a test sample, the presence of a modulator of zmse1 protein activity, comprising: culturing a cell into which has been introduced an expression vector as disclosed above, wherein the cell expresses the zmse1 protein encoded by the DNA segment in the presence and absence of a test sample; and comparing levels of activity of zmse1 in the presence and absence of a test sample, by a biological or biochemical assay; and determining from the comparison, the presence of modulator of zmse1 activity in the test sample.

In another aspect, the present invention provides a method for detecting a genetic abnormality in a patient, comprising: obtaining a genetic sample from a patient; producing a first reaction product by incubating the genetic sample with a polynucleotide comprising at least 14 contiguous nucleotides of SEQ ID NO:1 or the complement of SEQ ID NO:1, under conditions wherein said polynucleotide will hybridize to complementary polynucleotide sequence; visualizing the first reaction product; and comparing said first reaction product to a control reaction product from a wild type patient, wherein a difference between said first reaction product and said control reaction product is indicative of a genetic abnormality in the patient.

In another aspect, the present invention provides a method for detecting a cancer in a patient, comprising: obtaining a tissue or biological sample from a patient; incubating the tissue or biological sample with an antibody of claim 19 under conditions wherein the antibody binds to its complementary polypeptide in the tissue or biological sample; visualizing the antibody bound in the tissue or biological sample; and comparing levels of antibody bound in the tissue or biological sample from the patient to a normal control tissue or biological sample, wherein an increase or decrease in the level of antibody bound to the patient tissue or biological sample relative to the normal control tissue or biological sample is indicative of a cancer in the patient.

In another aspect, the present invention provides a method for detecting a cancer in a patient, comprising: obtaining a tissue or biological sample from a patient; labeling a polynucleotide comprising at least 14 contiguous nucleotides of SEQ

ID NO:1 or the complement of SEQ ID NO:1; incubating the tissue or biological sample with under conditions wherein the polynucleotide will hybridize to complementary polynucleotide sequence; visualizing the labeled polynucleotide in the tissue or biological sample; and comparing the level of labeled polynucleotide hybridization in the tissue or biological sample from the patient to a normal control tissue or biological sample, wherein an increase or decrease in the labeled polynucleotide hybridization to the patient tissue or biological sample relative to the normal control tissue or biological sample is indicative of a cancer in the patient.

Within another aspect, the present invention provides a transgenic mouse, wherein the mouse over-expresses residue 1 (Met) to residue 356 (Val) of SEQ ID NO:2) or residue 1 (Met) to residue 349 (Val) of SEQ ID NO:5. In one embodiment, the transgenic mouse disclosed above expresses residue 1 (Met) to residue 356 (Val) of SEQ ID NO:2) or residue 1 (Met) to residue 349 (Val) of SEQ ID NO:5 using a tissue-specific or tissue-restricted promoter. In another embodiment, the transgenic mouse disclosed above expresses residue 1 (Met) to residue 356 (Val) of SEQ ID NO:2) or residue 1 (Met) to residue 349 (Val) of SEQ ID NO:5 using an epithelial-specific, colon-specific, or ovary-specific promoter. In another embodiment, the transgenic mouse disclosed above does not expresses residue 1 (Met) to residue 349 (Val) of SEQ ID NO:5, relative to a normal mouse.

These and other aspects of the invention will become evident upon reference to the following detailed description of the invention and attached drawings.

Prior to setting forth the invention in detail, it may be helpful to the understanding thereof to define the following terms:

The term "affinity tag" is used herein to denote a polypeptide segment that can be attached to a second polypeptide to provide for purification or detection of the second polypeptide or provide sites for attachment of the second polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson et al., EMBO J. 4:1075, 1985; Nilsson et al., Methods Enzymol. 198:3, 1991), glutathione S transferase (Smith and Johnson, Gene 67:31, 1988), Glu-Glu affinity tag (Grussenmeyer et al., Proc. Natl. Acad. Sci. USA 82:7952-

4, 1985), substance P, Flag™ peptide (Hopp et al., Biotechnology 6:1204-10, 1988), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general, Ford et al., Protein Expression and Purification 2: 95-107, 1991. DNAs encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, NJ).

The term "allelic variant" is used herein to denote any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

The terms "amino-terminal" (also, "N-Terminal") and "carboxyl-terminal" (also "C-terminal") are used herein to denote positions within polypeptides. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a polypeptide is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete polypeptide.

The term "complement/anti-complement pair" denotes non-identical moieties that form a non-covalently associated, stable pair under appropriate conditions. For instance, biotin and avidin (or streptavidin) are prototypical members of a complement/anti-complement pair. Other exemplary complement/anti-complement pairs include receptor/ligand pairs, antibody/antigen (or hapten or epitope) pairs, sense/antisense polynucleotide pairs, and the like. Where subsequent dissociation of the complement/anti-complement pair is desirable, the complement/anti-complement pair preferably has a binding affinity of $<10^9$ M^{-1}.

The term "complements of a polynucleotide molecule" denotes a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATGCACGGG 3' is complementary to 5' CCCGTGCAT 3'.

The term "contig" denotes a polynucleotide that has a contiguous stretch of identical or complementary sequence to another polynucleotide. Contiguous sequences are said to "overlap" a given stretch of polynucleotide sequence either in their entirety or along a partial stretch of the polynucleotide. For example, representative contigs to the polynucleotide sequence 5'-ATGGAGCTT-3' are 5'-AGCTTgagt-3' and 3'-tcgacTACC-5'.

The term "degenerate nucleotide sequence" denotes a sequence of nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide). Degenerate codons contain different triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

The term "expression vector" is used to denote a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription. Such additional segments include promoter and terminator sequences, and may also include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

The term "isolated", when applied to a polynucleotide, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, Nature 316:774-78, 1985).

An "isolated" polypeptide or protein is a polypeptide or protein that is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated polypeptide is substantially free of other

polypeptides, particularly other polypeptides of animal origin. It is preferred to provide the polypeptides in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When used in this context, the term "isolated" does not exclude the presence of the same polypeptide in alternative physical forms, such as dimers or alternatively glycosylated or derivatized forms.

The term "operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert for their intended purposes, e.g., transcription initiates in the promoter and proceeds through the coding segment to the terminator.

The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

"Paralogs" are distinct but structurally related proteins made by an organism. Paralogs are believed to arise through gene duplication. For example, α-globin, β-globin, and myoglobin are paralogs of each other.

A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic molecules. Sizes of polynucleotides are expressed as base pairs (abbreviated "bp"), nucleotides ("nt"), or kilobases ("kb"). Where the context allows, the latter two terms may describe polynucleotides that are single-stranded or double-stranded. When the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term "base pairs". It will be recognized by those skilled in the art that the two strands of a double-stranded polynucleotide may differ slightly in length and that the ends thereof may be staggered as a result of enzymatic cleavage; thus all nucleotides within a double-stranded polynucleotide molecule may not be paired.

A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, whether produced naturally or synthetically. Polypeptides of less than about 10 amino acid residues are commonly referred to as "peptides".

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures; substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

The term "receptor" denotes a cell-associated protein that binds to a bioactive molecule (i.e., a ligand) and mediates the effect of the ligand on the cell. Membrane-bound receptors are characterized by a multi-peptide structure, for example, comprising an extracellular ligand-binding domain and an intracellular effector domain that is typically involved in signal transduction. Binding of ligand to receptor results in a conformational change in the receptor that causes an interaction between the effector domain and other molecule(s) in the cell. This interaction in turn leads to an alteration in the metabolism of the cell. Metabolic events that are linked to receptor-ligand interactions include gene transcription, phosphorylation, dephosphorylation, increases in cyclic AMP production, mobilization of cellular calcium, mobilization of membrane lipids, cell adhesion, hydrolysis of inositol lipids and hydrolysis of phospholipids. In general, receptors can be membrane bound, cytosolic or nuclear; monomeric (e.g., thyroid stimulating hormone receptor, beta-adrenergic receptor) or multimeric (e.g., PDGF receptor, growth hormone receptor, IL-3 receptor, GM-CSF receptor, G-CSF receptor, erythropoietin receptor and IL-6 receptor).

The term "secretory signal sequence" denotes a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

The term "splice variant" is used herein to denote alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Molecular weights and lengths of polymers determined by imprecise analytical methods (e.g., gel electrophoresis) will be understood to be approximate values. When such a value is expressed as "about" X or "approximately" X, the stated value of X will be understood to be accurate to ±10%.

All references cited herein are incorporated by reference in their entirety.

The present invention is based in part upon the discovery of a novel DNA sequence that encodes a protein having a CRIB motif (Burbelo, P.D. et al., J. Biol. Chem. 270:29071-29074, 1995). The polypeptide has been designated zmse1. The novel zmse1 was found and its corresponding cDNA was sequenced. The novel polypeptide encoded by the cDNA showed limited homology with MSE55 (Bahou, W.F., et al., J. Biol. Chem. 267:13986-13992, 1994; Burbelo, P.D. et al., Proc. Natl. Acad. Sci. USA 96:9083-9088, 1999). The zmse1 polynucleotide sequence encodes the entire coding sequence of the predicted protein. Zmse1 is a novel protein that may be involved in regulating actin polymerization and resulting structures, cytoskeletal organization, proliferation, cell transformation, motility, cell invasion, metastasis, transport or secretion, tissue contractility, involved in an apoptotic cellular pathway, or the like.

The sequence of the zmse1 polypeptide was deduced from a single clone that contained its corresponding polynucleotide sequence. The clone was obtained from a human K562 cell (ATCC Cat. No. CCL 243) library. Other libraries that might also be searched for such sequences include tumor cell and tissue libraries PBLs, testis, gastrointestinal, prostate, lung, adrenal gland, and the like.

The nucleotide sequence of a representative human zmse1-encoding DNA is described in SEQ ID NO:1, and its deduced 356 residue amino acid sequence is described in SEQ ID NO:2. In its entirety, the human zmse1 polypeptide (SEQ ID NO:2) represents a full-length polypeptide segment (residue 1 (Met) to 356 (Val) of SEQ ID NO:2). The domains and structural features of the zmse1 polypeptide are further described below.

Analysis of the zmse1 polypeptide encoded by the DNA sequence of SEQ ID NO:1 revealed an open reading frame encoding 356 amino acids (SEQ ID NO:2) comprising a mature polypeptide. Zmse1 contains a CRIB motif (SEQ ID NO:6) comprising amino acid residue number 27 (Ile) to amino acid residue number 41 (Gly) of SEQ ID NO:2. This CRIB motif is conserved and identical in both the human and murine forms of zmse1 (see, Figure 1; and, amino acid residue number 27 (Ile) to amino acid residue number 41 (Gly) of SEQ ID NO:2 and SEQ ID NO:5), suggesting that a minor sequence difference therein could affect the binding of this effector with its target. For example, amino acid mutations in the CRIB motif of MSE55 and a single point mutation in the CRIB motif of WASP decrease or abolish Cdc42 binding and hence the biological activity associated therewith (See, Burbelo, P.D. et al., Proc. Natl. Acad. Sci. USA supra.; and Miki, H. et al., Nature 391:93-96, 1998). However, sequences outside the CRIB motif are also important for the activity of these effector proteins (See, Zohn, I.M. et al., Oncogene 17:1415-1438, 1998). Moreover, zmse1 contains a highly conserved N-terminal domain of approximately 150 amino acid residues (residues 1 (Met) to 147 (Ala) of SEQ ID NO:2; and residues 1 (Met) to 145 (Ala) of SEQ ID NO:5); and a more variable C-terminal domain of approximately 180 amino acid residues compressing residues 148 (Asn) to 336 (Ser) of SEQ ID NO:2, and residues 146 (Asp) to 329 (Pro) of SEQ ID NO:5); and a highly conserved C-terminal tail comprising residues 337 (Arg) to 356 (Val) of SEQ ID NO:2, and 329 (Arg) to 350 (Val) of SEQ ID NO:5. Moreover zmse1 contains several phosphorylation sites that are conserved between the human and mouse polypeptides, and are shown in SEQ ID NO:2 as follows: Ser_{295}-Ala_{296}-Arg_{297}; Ser_{86}-Lys_{87}-Arg_{88}; Ser_{14}-Lys_{15}-Arg_{16}; Ser_{105}-Leu_{106}-Arg_{107}; Ser_{80}-Arg_{81}-Lys_{82}; and Thr_{303}-Thr_{304}-Arg_{305}. The corresponding mouse phosphorylation sites can be determined with reference to Figure 2 and SEQ ID NO:5.

As zmse1 is likely involved in signal transduction, some or all of these phosphorylation sites may be essential for zmse1 activity. Those skilled in the art will recognize that these domain boundaries are approximate, and are based on alignments with known proteins and predictions of protein folding. The corresponding polynucleotides encoding the zmse1 polypeptide regions, domains, motifs, residues and sequences described above are as shown in SEQ ID NO:1 (human zmse1) and SEQ ID NO:4 (mouse zmse1).

The presence of conserved motifs, such as the CRIB motif, and low variance motifs generally correlates with or defines important structural regions in proteins. Regions of low variance (e.g., hydrophobic clusters) are generally present in regions of structural importance (Sheppard, P. et al., supra.). Such regions of low variance often contain rare or infrequent amino acids, such as Tryptophan. The regions flanking and between such conserved and low variance motifs may be more variable, but are often functionally significant because they may relate to or define important structures and activities such as guanosine nucleotide binding domains, activation domains, biological and enzymatic activity, signal transduction, cell-cell interaction, tissue or intracellular localization domains and the like. For example, alignment of zmse1 with related polypeptides, for example MSE55, and the presence of a conserved CRIB motif supports that the correlating structural and functional domains of zmse1 are significant in determining that zmse1 is a Rho family effector.

The regions of conserved amino acid residues in zmse1, described above, can be used as tools to identify new family members. For instance, reverse transcription-polymerase chain reaction (RT-PCR) can be used to amplify sequences encoding the conserved regions from RNA obtained from a variety of tissue sources or cell lines. In particular, highly degenerate primers designed from the zmse1 sequences are useful for this purpose. Designing and using such degenerate primers may be readily performed by one of skill in the art.

The present invention also provides polynucleotide molecules, including DNA and RNA molecules, that encode the zmse1 polypeptides disclosed herein. Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among these polynucleotide

molecules. SEQ ID NO:3 is a degenerate DNA sequence that encompasses all DNAs that encode the human zmse1 polypeptide of SEQ ID NO:2. Those skilled in the art will recognize that the degenerate sequence of SEQ ID NO:3 also provides all RNA sequences encoding SEQ ID NO:2 by substituting U for T. Thus, zmse1 polypeptide-encoding polynucleotides comprising nucleotide 1 to nucleotide 1068 of SEQ ID NO:3 and their RNA equivalents are contemplated by the present invention. Table 1 sets forth the one-letter codes used within SEQ ID NO:3 to denote degenerate nucleotide positions. "Resolutions" are the nucleotides denoted by a code letter. "Complement" indicates the code for the complementary nucleotide(s). For example, the code Y denotes either C or T, and its complement R denotes A or G, A being complementary to T, and G being complementary to C.

TABLE I

Nucleotide	Resolution	Complement	Resolution
A	A	T	T
C	C	G	G
G	G	C	C
T	T	A	A
R	A|G	Y	C|T
Y	C|T	R	A|G
M	A|C	K	G|T
K	G|T	M	A|C
S	C|G	S	C|G
W	A|T	W	A|T
H	A|C|T	D	A|G|T
B	C|G|T	V	A|C|G
V	A|C|G	B	C|G|T
D	A|G|T	H	A|C|T
N	A|C|G|T	N	A|C|G|T

The degenerate codons used in SEQ ID NO:3, encompassing all possible codons for a given amino acid, are set forth in Table 2.

TABLE 2

Amino Acid	One Letter Code	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	ACN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAY
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	.	TAA TAG TGA	TRR
Asn|Asp	B		RAY
Glu|Gln	Z		SAR
Any	X		NNN

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship exists between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequence of SEQ ID NO:2. Variant sequences can be readily tested for functionality as described herein.

One of ordinary skill in the art will also appreciate that different species can exhibit "preferential codon usage." In general, see, Grantham, et al., Nuc. Acids Res. 8:1893-912, 1980; Haas, et al. Curr. Biol. 6:315-24, 1996; Wain-Hobson, et al., Gene 13:355-64, 1981; Grosjean and Fiers, Gene 18:199-209, 1982; Holm, Nuc. Acids Res. 14:3075-87, 1986; Ikemura, J. Mol. Biol. 158:573-97, 1982. As used herein, the term "preferential codon usage" or "preferential codons" is a term of art referring to protein translation codons that are most frequently used in cells of a certain species, thus favoring one or a few representatives of the possible codons encoding each amino acid (See Table 2). For example, the amino acid Threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian cells ACC is the most commonly used codon; in other species, for example, insect cells, yeast, viruses or bacteria, different Thr codons may be preferential. Preferential codons for a particular species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species. Therefore, the degenerate codon sequence disclosed in SEQ ID NO:3 serves as a template for optimizing expression of polynucleotides in various cell types and species commonly used in the art and disclosed herein. Sequences containing preferential codons can be tested and optimized for expression in various species, and tested for functionality as disclosed herein.

Within preferred embodiments of the invention the isolated polynucleotides will hybridize to similar sized regions of SEQ ID NO:1, or a sequence complementary thereto, under stringent conditions. In general, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Numerous equations for calculating T_m are known in the art, and are specific for DNA, RNA and DNA-RNA hybrids and polynucleotide probe sequences of varying length (see, for example, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor Press 1989); Ausubel *et al.*, (eds.), Current Protocols in Molecular Biology (John Wiley and Sons, Inc. 1987); Berger and Kimmel (eds.), Guide to Molecular Cloning Techniques, (Academic Press, Inc. 1987); and Wetmur, Crit. Rev. Biochem. Mol. Biol. 26:227, 1990). Sequence analysis software such as OLIGO 6.0 (LSR; Long Lake, MN) and *Primer Premier 4.0* (Premier Biosoft International; Palo Alto, CA), as well as sites on the Internet, are available tools for analyzing a given sequence and calculating T_m based on user defined criteria. Such programs can also analyze a given sequence under defined conditions and identify suitable probe sequences. Typically, hybridization of longer polynucleotide sequences (e.g., >50 base pairs) is performed at temperatures of about 20-25°C below the calculated T_m. For smaller probes (e.g., <50 base pairs) hybridization is typically carried out at the T_m or 5-10°C below. This allows for the maximum rate of hybridization for DNA-DNA and DNA-RNA hybrids. Higher degrees of stringency at lower temperatures can be achieved with the addition of formamide which reduces the T_m of the hybrid about 1°C for each 1% formamide in the buffer solution. Suitable stringent hybridization conditions are equivalent to about a 5 h to overnight incubation at about 42°C in a solution comprising: about 40-50% formamide, up to about 6X SSC, about 5X Denhardt's solution, zero up to about 10% dextran sulfate, and about 10-20 µg/ml denatured commercially-available carrier DNA. Generally, such stringent conditions include temperatures of 20-70°C and a hybridization buffer containing up to 6x SSC and 0-50% formamide; hybridization is then followed by washing filters in up to about 2X SSC. For example, a suitable wash stringency is equivalent to 0.1X SSC to

2X SSC, 0.1% SDS, at 55°C to 65°C. Different degrees of stringency can be used during hybridization and washing to achieve maximum specific binding to the target sequence. Typically, the washes following hybridization are performed at increasing degrees of stringency to remove non-hybridized polynucleotide probes from hybridized complexes. Stringent hybridization and wash conditions depend on the length of the probe, reflected in the Tm, hybridization and wash solutions used, and are routinely determined empirically and experimentally by one of skill in the art.

As previously noted, the isolated polynucleotides of the present invention include DNA and RNA. Methods for preparing DNA and RNA are well known in the art. In general, RNA is isolated from a tissue or cell that produces large amounts of zmse1 RNA. Such tissues and cells are identified by Northern blotting (Thomas, Proc. Natl. Acad. Sci. USA 77:5201, 1980), and include intestinal tissues, prostate, ovary, testis, spleen, pancreas, heart, skeletal muscle and the like. Total RNA can be prepared using guanidinium isothiocyanate extraction followed by isolation by centrifugation in a CsCl gradient (Chirgwin et al., Biochemistry 18:52-94, 1979). Poly $(A)^+$ RNA is prepared from total RNA using the method of Aviv and Leder (Proc. Natl. Acad. Sci. USA 69:1408-12, 1972). Complementary DNA (cDNA) is prepared from poly$(A)^+$ RNA using known methods. In the alternative, genomic DNA can be isolated. Polynucleotides encoding zmse1 polypeptides are then identified and isolated by, for example, hybridization or PCR.

A full-length clone encoding zmse1 can be obtained by conventional cloning procedures. Complementary DNA (cDNA) clones are preferred, although for some applications (e.g., expression in transgenic animals) it may be preferable to use a genomic clone, or to modify a cDNA clone to include at least one genomic intron. Methods for preparing cDNA and genomic clones are well known and within the level of ordinary skill in the art, and include the use of the sequence disclosed herein, or parts thereof, for probing or priming a library. Expression libraries can be probed with antibodies to zmse1, ligand fragments, or other specific binding partners.

Zmse1 polynucleotide sequences disclosed herein can also be used as probes or primers to clone 5' non-coding regions of a zmse1 gene. In view of the tissue-specific expression observed for zmse1 by Northern blotting, this gene region

may provide for expression in many cell and tissue types. Promoter elements from a zmse1 gene could thus be used to direct the ubiquitous expression of heterologous genes in, for example, transgenic animals or patients treated with gene therapy. Cloning of 5' flanking sequences also facilitates production of zmse1 proteins by "gene activation" as disclosed in U.S. Patent No. 5,641,670. Briefly, expression of an endogenous zmse1 gene in a cell is altered by introducing into the zmse1 locus a DNA construct comprising at least a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The targeting sequence is a zmse1 5' non-coding sequence that permits homologous recombination of the construct with the endogenous zmse1 locus, whereby the sequences within the construct become operably linked with the endogenous zmse1 coding sequence. In this way, an endogenous zmse1 promoter can be replaced or supplemented with other regulatory sequences to provide enhanced, tissue-specific, or otherwise regulated expression.

The polynucleotides of the present invention can also be synthesized using DNA synthesis machines. If chemically synthesized double stranded DNA is required for an application such as the synthesis of a DNA or a DNA fragment, then each complementary strand is made separately, for example via the phosphoramidite method known in the art. The production of short polynucleotides (60 to 80 bp) is technically straightforward and can be accomplished by synthesizing the complementary strands and then annealing them. However, for producing longer polynucleotides (longer than about 300 bp), special strategies are usually employed. For example, synthetic DNAs (double-stranded) are assembled in modular form from single-stranded fragments that are from 20 to 100 nucleotides in length. One method for building a synthetic DNA involves producing a set of overlapping, complementary oligonucleotides. Each internal section of the DNA has complementary 3' and 5' terminal extensions designed to base pair precisely with an adjacent section. After the DNA is assembled, the process is completed by ligating the nicks along the backbones of the two strands. In addition to the protein coding sequence, synthetic DNAs can be designed with terminal sequences that facilitate insertion into a restriction endonuclease site of a cloning vector. Alternative ways to prepare a full-length DNA are also known in the art. See Glick and Pasternak, Molecular Biotechnology, Principles &

Applications of Recombinant DNA, (ASM Press, Washington, D.C. 1994); Itakura et al., Annu. Rev. Biochem. 53: 323-56, 1984 and Climie et al., Proc. Natl. Acad. Sci. USA 87:633-7, 1990.

The present invention further provides counterpart polypeptides and polynucleotides from other species (orthologs). These species include, but are not limited to mammalian, avian, amphibian, reptile, fish, insect and other vertebrate and invertebrate species. Of particular interest are zmse1 polypeptides from other mammalian species, including murine, porcine, ovine, bovine, canine, feline, equine, and other primate polypeptides. Orthologs of human zmse1 can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses zmse1 as disclosed herein. Suitable sources of mRNA can be identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line. A zmse1-encoding cDNA can then be isolated by a variety of methods, such as by probing with a complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be cloned using the polymerase chain reaction, or PCR (Mullis, U.S. Patent No. 4,683,202), using primers designed from the representative human zmse1 sequence disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to zmse1 polypeptide. Similar techniques can also be applied to the isolation of genomic clones.

A polynucleotide sequence for the mouse ortholog of human zmse1 has been identified and cloned and is shown in SEQ ID NO:4 and the corresponding amino acid sequence shown in SEQ ID NO:5. Analysis of the mouse zmse1 polypeptide encoded by the DNA sequence of SEQ ID NO:4 revealed an open reading frame encoding 349 amino acids (SEQ ID NO:5) comprising a CRIB motif, N-terminal domain, C-terminal domain, and C-terminal tail as described above. A comparison of the human and mouse amino acid sequences reveals that both the human and orthologous polypeptides contain corresponding structural features described above

(See, Figure 2). There is about 78% identity between the mouse and human amino acid sequences over the entire amino acid sequence corresponding to SEQ ID NO:2 and SEQ ID NO:5. There is about 100% identity over the CRIB motif corresponding to amino acid residue 27 (Ile) to amino acid residue 41 (Gly) of SEQ ID NO:2 and SEQ ID NO:5. There is about 91% identity a between the mouse and human zmse1 sequences over the conserved N-terminal domain corresponding to residues 1 (Met) to 147 (Ala) of SEQ ID NO:2; and residues 1 (Met) to 145 (Ala) of SEQ ID NO:5. There is about 73% identity a between the mouse and human zmse1 sequences over the variable C-terminal domain corresponding to residues 148 (Asn) to 336 (Ser) of SEQ ID NO:2, and residues 146 (Asp) to 329 (Pro) of SEQ ID NO:5). There is about 95% identity a between the mouse and human zmse1 sequences over the conserved C-terminal tail corresponding to residues 337 (Arg) to 356 (Val) of SEQ ID NO:2, and 330 (Arg) to 349 (Val) of SEQ ID NO:5. The above percent identities were determined using a FASTA program with ktup=1, gap opening penalty=12, gap extension penalty=2, and substitution matrix=BLOSUM62, with other parameters set as default. The corresponding polynucleotides encoding the mouse zmse1 polypeptide regions, domains, motifs, residues and sequences described above are as shown in SEQ ID NO:4.

Those skilled in the art will recognize that the sequence disclosed in SEQ ID NO:1 represents a single allele of human zmse1 and that allelic variation and alternative splicing are expected to occur. Allelic variants of this sequence can be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO:1, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants of SEQ ID NO:2. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the zmse1 polypeptide are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individuals or tissues according to standard procedures known in the art.

The present invention also provides isolated zmsc1 polypeptides that are substantially similar to the polypeptides of SEQ ID NO:2 and their orthologs. The term "substantially similar" is used herein to denote polypeptides having 70%, preferably 80%, more preferably at least 85%, sequence identity to the sequences shown in SEQ ID NO:2 or their orthologs. Such polypeptides will more preferably be at least 90% identical, and most preferably 95% or more identical to SEQ ID NO:2 or its orthologs.) Percent sequence identity is determined by conventional methods. See, for example, Altschul et al., Bull. Math. Bio. 48: 603-16, 1986 and Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-9, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix of Henikoff and Henikoff (ibid.) as shown in Table 3 (amino acids are indicated by the standard one-letter codes). The percent identity is then calculated as:

$$\frac{\text{Total number of identical matches}}{\left[\begin{array}{l}\text{length of the longer sequence plus the}\\ \text{number of gaps introduced into the longer}\\ \text{sequence in order to align the two sequences}\end{array}\right]} \times 100$$

Table 3

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-3	-2	-1	-2	-3	-3	-2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Sequence identity of polynucleotide molecules is determined by similar methods using a ratio as disclosed above.

Those skilled in the art appreciate that there are many established algorithms available to align two amino acid sequences. The "FASTA" similarity search algorithm of Pearson and Lipman is a suitable protein alignment method for examining the level of identity shared by an amino acid sequence disclosed herein and the amino acid sequence of a putative variant zmsc1. The FASTA algorithm is described by Pearson and Lipman, Proc. Nat'l Acad. Sci. USA 85:2444, 1988; and by Pearson, Meth. Enzymol. 183:63, 1990.

Briefly, FASTA first characterizes sequence similarity by identifying regions shared by the query sequence (*e.g.*, SEQ ID NO:2) and a test sequence that have either the highest density of identities (if the ktup variable is 1) or pairs of identities (if ktup=2), without considering conservative amino acid substitutions, insertions, or deletions. The ten regions with the highest density of identities are then rescored by comparing the similarity of all paired amino acids using an amino acid substitution matrix, and the ends of the regions are "trimmed" to include only those residues that contribute to the highest score. If there are several regions with scores greater than the "cutoff" value (calculated by a predetermined formula based upon the length of the sequence and the ktup value), then the trimmed initial regions are examined to determine whether the regions can be joined to form an approximate alignment with gaps. Finally, the highest scoring regions of the two amino acid sequences are aligned using a modification of the Needleman-Wunsch-Sellers algorithm (Needleman and Wunsch, J. Mol. Biol. 48:444, 1970; Sellers, SIAM J. Appl. Math. 26:787, 1974), which allows for amino acid insertions and deletions. Preferred parameters for FASTA analysis are: ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62. These parameters can be introduced into a FASTA program by modifying the scoring matrix file ("SMATRIX"), as explained in Appendix 2 of Pearson, Meth. Enzymol., supra.

FASTA can also be used to determine the sequence identity of nucleic acid molecules using a ratio as disclosed above. For nucleotide sequence comparisons,

the ktup value can range between one to six, preferably from three to six, most preferably three, with other FASTA program parameters set as default.

The BLOSUM62 table (Table 3) is an amino acid substitution matrix derived from about 2,000 local multiple alignments of protein sequence segments, representing highly conserved regions of more than 500 groups of related proteins (Henikoff and Henikoff, Proc. Nat'l Acad. Sci. USA 89:10915, 1992). Accordingly, the BLOSUM62 substitution frequencies can be used to define conservative amino acid substitutions that may be introduced into the amino acid sequences of the present invention. Although it is possible to design amino acid substitutions based solely upon chemical properties (as discussed below), the language "conservative amino acid substitution" preferably refers to a substitution represented by a BLOSUM62 value of greater than -1. For example, an amino acid substitution is conservative if the substitution is characterized by a BLOSUM62 value of 0, 1, 2, or 3. According to this system, preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 1 (e.g., 1, 2 or 3), while more preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 2 (e.g., 2 or 3).

Variant zmse1 polypeptides or substantially similar zmse1 polypeptides are characterized as having one or more amino acid substitutions, deletions or additions. These changes are preferably of a minor nature, that is conservative amino acid substitutions (see Table 4) and other substitutions that do not significantly affect the folding or activity of the polypeptide; small deletions, typically of one to about 30 amino acids; and amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues, or an affinity tag. The present invention thus includes polypeptides of from about 330 to about 385 amino acid residues that comprise a sequence that is at least 90%, preferably at least 95%, and more preferably 99% or more identical to the corresponding region of SEQ ID NO:2. Polypeptides comprising affinity tags can further comprise a proteolytic cleavage site between the zmse1 polypeptide and the affinity tag. Preferred such sites include thrombin cleavage sites and factor Xa cleavage sites.

Table 4

Conservative amino acid substitutions

Basic:	arginine
	lysine
	histidine
Acidic:	glutamic acid
	aspartic acid
Polar:	glutamine
	asparagine
Hydrophobic:	leucine
	isoleucine
	valine
Aromatic:	phenylalanine
	tryptophan
	tyrosine
Small:	glycine
	alanine
	serine
	threonine
	methionine

The present invention further provides a variety of other polypeptide fusions and related multimeric proteins comprising one or more polypeptide fusions. For example, a zmsel polypeptide can be prepared as a fusion to a dimerizing protein as disclosed in U.S. Patents Nos. 5,155,027 and 5,567,584. Preferred dimerizing proteins in this regard include immunoglobulin constant region domains. Immunoglobulin-zmsel polypeptide fusions can be expressed in genetically engineered cells to produce a variety of multimeric zmsel analogs. Auxiliary domains can be fused to zmsel polypeptides to target them to specific cells, tissues, or macromolecules (e.g., collagen). For example, a zmsel polypeptide or protein could be targeted to a

predetermined cell type by fusing a zmsel polypeptide to a ligand that specifically binds to a receptor on the surface of the target cell. In this way, polypeptides and proteins can be targeted for therapeutic or diagnostic purposes. A zmsel polypeptide can be fused to two or more moieties, such as an affinity tag for purification and a targeting domain. Polypeptide fusions can also comprise one or more cleavage sites, particularly between domains. See, Tuan et al., Connective Tissue Research 34:1-9, 1996.

The proteins of the present invention can also comprise non-naturally occurring amino acid residues. Non-naturally occurring amino acids include, without limitation, *trans*-3-methylproline, 2,4-methanoproline, *cis*-4-hydroxyproline, *trans*-4-hydroxyproline, *N*-methylglycine, *allo*-threonine, methylthreonine, hydroxyethylcysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, 3,3-dimethylproline, *tert*-leucine, norvaline, 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an *in vitro* system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell-free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., J. Am. Chem. Soc. 113:2722, 1991; Ellman et al., Methods Enzymol. 202:301, 1991; Chung et al., Science 259:806-9, 1993; and Chung et al., Proc. Natl. Acad. Sci. USA 90:10145-9, 1993). In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., J. Biol. Chem. 271:19991-8, 1996). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the protein in place of its natural counterpart.

See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by *in vitro* chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for zinsel amino acid residues.

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244: 1081-5, 1989; Bass et al., Proc. Natl. Acad. Sci. USA 88:4498-502, 1991). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity as disclosed below to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton et al., J. Biol. Chem. 271:4699-708, 1996. Sites of ligand-receptor or other biochemical interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., Science 255:306-12, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of essential amino acids can also be inferred from analysis of homologies with related phosphodiesterases.

Determining amino acid residues that are within regions or domains that are critical to maintaining structural integrity is within the skill of one in the art. Within these regions one can determine specific residues that will be more or less tolerant of change and maintain the overall tertiary structure of the molecule. Methods for analyzing sequence structure include, but are not limited to, alignment of multiple sequences with high amino acid or nucleotide identity and computer analysis using available software (e.g., the Insight II® viewer and homology modeling tools; MSI, San Diego, CA), secondary structure propensities, binary patterns, complementary packing and buried polar interactions (Barton, Current Opin. Struct. Biol. 5:372-376, 1995 and

Cordes et al., Current Opin. Struct. Biol. 6:3-10, 1996). In general, when designing modifications to molecules or identifying specific fragments determination of structure will be accompanied by evaluating activity of modified molecules.

Amino acid sequence changes are made in zmse1 polypeptides so as to minimize disruption of higher order structure essential to biological activity. For example, when the zmse1 polypeptide comprises one or more helices, changes in amino acid residues will be made so as not to disrupt the helix geometry and other components of the molecule where changes in conformation abate some critical function, for example, binding of the molecule to its binding partners, or enzymatic function. The effects of amino acid sequence changes can be predicted by, for example, computer modeling as disclosed above or determined by analysis of crystal structure (see, e.g., Lapthorn et al., Nat. Struct. Biol. 2:266-268, 1995). Other techniques that are well known in the art compare folding of a variant protein to a standard molecule (e.g., the native protein). For example, comparison of the cysteine pattern in a variant and standard molecules can be made. Mass spectrometry and chemical modification using reduction and alkylation provide methods for determining cysteine residues which are associated with disulfide bonds or are free of such associations (Bean et al., Anal. Biochem. 201:216-226, 1992; Gray, Protein Sci. 2:1732-1748, 1993; and Patterson et al., Anal. Chem. 66:3727-3732, 1994). It is generally believed that if a modified molecule does not have the same disulfide bonding pattern as the standard molecule folding would be affected. Another well known and accepted method for measuring folding is circular dichroism (CD). Measuring and comparing the CD spectra generated by a modified molecule and standard molecule is routine (Johnson, Proteins 7:205-214, 1990). Crystallography is another well known method for analyzing folding and structure. Nuclear magnetic resonance (NMR), digestive peptide mapping and epitope mapping are also known methods for analyzing folding and structural similarities between proteins and polypeptides (Schaanan et al., Science 257:961-964, 1992).

A Hopp/Woods hydrophilicity profile of the zmse1 protein sequence as shown in SEQ ID NO:2 can be generated (Hopp et al., Proc. Natl. Acad. Sci.78:3824-3828, 1981; Hopp, J. Immun. Meth. 88:1-18, 1986 and Triquier et al., Protein

Engineering 11:153-169, 1998). The profile is based on a sliding six-residue window. Buried G, S, and T residues and exposed H, Y, and W residues were ignored (See, Figure 1). For example, in zmse1, hydrophilic regions include: (1) amino acid number 96 (Glu) to amino acid number 101 (Asp) of SEQ ID NO:2; (2) amino acid number 226 (Asp) to amino acid number 231 (Asp) of SEQ ID NO:2; (3) amino acid number 346 (Met) to amino acid number 351 (Glu) of SEQ ID NO:2; (4) amino acid number 347 (Asp) to amino acid number 352 (Asp) of SEQ ID NO:2; and (5) amino acid number 348 (Glu) to amino acid number 353 (Glu) of SEQ ID NO:2.

Those skilled in the art will recognize that hydrophilicity or hydrophobicity is taken into account when designing modifications in the amino acid sequence of a zmse1 polypeptide, so as not to disrupt the overall structural and biological profile. Of particular interest for replacement are hydrophobic residues selected from the group consisting of Val, Leu and Ile or the group consisting of Met, Gly, Ser, Ala, Tyr and Trp; for example, residues tolerant of substitution could include such residues as shown in SEQ ID NO: 2. Cysteine residues will be relatively intolerant of substitution.

The identities of essential amino acids can also be inferred from analysis of sequence similarity between known phosphodiesterase family members with zmse1. Using methods such as "FASTA" analysis described previously, regions of high similarity are identified within a family of proteins and used to analyze amino acid sequence for conserved regions. An alternative approach to identifying a variant zmse1 polynucleotide on the basis of structure is to determine whether a nucleic acid molecule encoding a potential variant zmse1 polynucleotide can hybridize to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, as discussed above.

Other methods of identifying essential amino acids in the polypeptides of the present invention are procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244:1081, 1989; Bass et al., Proc. Natl Acad. Sci. USA 88:4498, 1991; Coombs and Corey, "Site-Directed Mutagenesis and Protein Engineering," in Proteins: Analysis and Design, Angeletti (ed.), Academic Press, Inc., pp. 259-311, 1998). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and

the resultant mutant molecules are tested for biological activity as disclosed below to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton et al., J. Biol. Chem. 271:4699, 1996.

The present invention also includes functional fragments of zmse1 polypeptides and nucleic acid molecules encoding such functional fragments. A "functional" zmse1 or fragment thereof defined herein is characterized by its proliferative or differentiating activity, by its ability to induce or inhibit specialized cell functions, or by its ability to bind specifically to an anti-zmse1 antibody or zmse1 substrate or binding partner (either soluble or immobilized).

Routine deletion analyses of nucleic acid molecules can be performed to obtain functional fragments of a nucleic acid molecule that encodes a zmse1 polypeptide. As an illustration, DNA molecules having the nucleotide sequence of SEQ ID NO:1 or fragments thereof, can be digested with *Bal*31 nuclease to obtain a series of nested deletions. These DNA fragments are then inserted into expression vectors in proper reading frame, and the expressed polypeptides are isolated and tested for zmse1 activity, or for the ability to bind anti-zmse1 antibodies or zmse1 substrate or binding partner. One alternative to exonuclease digestion is to use oligonucleotide-directed mutagenesis to introduce deletions or stop codons to specify production of a desired zmse1 fragment. Alternatively, particular fragments of a zmse1 polynucleotide can be synthesized using the polymerase chain reaction.

Standard methods for identifying functional domains are well-known to those of skill in the art. For example, studies on the truncation at either or both termini of interferons have been summarized by Horisberger and Di Marco, Pharmac. Ther. 66:507, 1995. Moreover, standard techniques for functional analysis of proteins are described by, for example, Treuter et al., Molec. Gen. Genet. 240:113, 1993; Content et al., "Expression and preliminary deletion analysis of the 42 kDa 2-5A synthetase induced by human interferon," in Biological Interferon Systems, Proceedings of ISIR-TNO Meeting on Interferon Systems, Cantell (ed.), Nijhoff, pp. 65-72, 1987; Herschman, "The EGF Receptor," in Control of Animal Cell Proliferation 1, Boynton et al. (eds.), Academic Press, pp. 169-199, 1985; Coumailleau et al., J. Biol. Chem.

270:29270, 1995; Fukunaga et al., J. Biol. Chem. 270:25291, 1995; Yamaguchi et al., Biochem. Pharmacol. 50:1295, 1995; and Meisel et al., Plant Molec. Biol. 30:1, 1996.

In addition, the proteins of the present invention (or polypeptide fragments thereof) can be joined to other bioactive molecules, particularly other CRIB proteins or Rho effectors, to provide multi-functional molecules. For example, one or more domains or sub-fragments from zmse1 can be joined to other CRIB proteins to enhance their biological properties or efficiency of production.

The present invention thus provides a series of novel, hybrid molecules in which a segment comprising one or more of the domains or motifs of zmse1 is fused to another polypeptide. Fusion is preferably done by splicing at the DNA level to allow expression of chimeric molecules in recombinant production systems. The resultant molecules are then assayed for such properties as improved solubility, improved stability, prolonged clearance half-life, improved expression and secretion levels, and pharmacodynamics. Such hybrid molecules may further comprise additional amino acid residues (e.g. a polypeptide linker) between the component proteins or polypeptides.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-7, 1988) or Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-6, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30:10832-7, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

Variants of the disclosed zmse1 DNA and polypeptide sequences can be generated through DNA shuffling as disclosed by Stemmer, Nature 370:389-91, 1994, Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-51, 1994 and WIPO Publication WO 97/20078. Briefly, variant DNAs are generated by *in vitro* homologous recombination

by random fragmentation of a parent DNA followed by reassembly using PCR, resulting in randomly introduced point mutations. This technique can be modified by using a family of parent DNAs, such as allelic variants or DNAs from different species, to introduce additional variability into the process. Selection or screening for the desired activity, followed by additional iterations of mutagenesis and assay provides for rapid "evolution" of sequences by selecting for desirable mutations while simultaneously selecting against detrimental changes.

Mutagenesis methods as disclosed herein can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides in host cells. Mutagenized DNA molecules that encode active polypeptides (e.g., those with CRIB protein activity, that induce signal transduction, that exhibit Cdc42/Rac or other Rho protein binding, that bind anti-zmse1 antibodies, and the like) can be recovered from the host cells and rapidly sequenced. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

Using the methods discussed herein, one of ordinary skill in the art can identify and/or prepare a variety of polypeptide fragments or variants of SEQ ID NO:2 or that retain, for example, CRIB protein-like properties, protein binding activity, cytoskeletal rearranging activity, proliferation, induction of actin polymerization, cell motility or invasion, cell-cell communication, or signal transduction activity of the wild-type zmse1 protein. For example, using the methods described herein, one could identify a substrate binding domain in addition to the CRIB domain on zmse1; heterodimeric and homodimeric binding domains; guanosine nucleotide binding domains; other enzymatically active domains; other functional or structural domains; or other domains important for protein-protein interactions, biological activity, or signal transduction. Such polypeptides may also include additional polypeptide segments, such as affinity tags, as generally disclosed herein.

For any zmse1 polypeptide, including variants and fusion proteins, one of ordinary skill in the art can readily generate a fully degenerate polynucleotide sequence encoding that variant using the information set forth in Tables 1 and 2 above.

The zmse1 polypeptides of the present invention, including full-length polypeptides, biologically active fragments, and fusion polypeptides, can be produced in genetically engineered host cells according to conventional techniques. Suitable host cells are those cell types that can be transformed or transfected with exogenous DNA and grown in culture, and include bacteria, fungal cells, and cultured higher eukaryotic cells. Eukaryotic cells, particularly cultured cells of multicellular organisms, are preferred. Techniques for manipulating cloned DNA molecules and introducing exogenous DNA into a variety of host cells are disclosed by Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and Ausubel et al., eds., Current Protocols in Molecular Biology, John Wiley and Sons, Inc., NY, 1987.

In general, a DNA sequence encoding a zmse1 polypeptide is operably linked to other genetic elements required for its expression, generally including a transcription promoter and terminator, within an expression vector. The vector will also commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers may be provided on separate vectors, and replication of the exogenous DNA may be provided by integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a matter of routine design within the level of ordinary skill in the art. Many such elements are described in the literature and are available through commercial suppliers.

To direct a zmse1 polypeptide into the secretory pathway of a host cell, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) is provided in the expression vector. The secretory signal sequence may be derived from any secreted protein (e.g., t-PA) or synthesized *de novo*. The secretory signal sequence is operably linked to the zmse1 DNA sequence, i.e., the two sequences are joined in the correct reading frame and positioned to direct the newly synthesized polypeptide into the secretory pathway of the host cell. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the polypeptide of interest, although certain secretory signal sequences may be positioned elsewhere in the DNA

sequence of interest (see, e.g., Welch et al., U.S. Patent No. 5,037,743; Holland et al., U.S. Patent No. 5,143,830).

Cultured mammalian cells are suitable hosts within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., Cell 14:725, 1978; Corsaro and Pearson, Somatic Cell Genetics 7:603, 1981; Graham and Van der Eb, Virology 52:456, 1973), electroporation (Neumann et al., EMBO J. 1:841-5, 1982), DEAE-dextran mediated transfection (Ausubel et al., ibid.), and liposome-mediated transfection (Hawley-Nelson et al., Focus 15:73, 1993; Ciccarone et al., Focus 15:80, 1993, and viral vectors (Miller and Rosman, BioTechniques 7:980-90, 1989; Wang and Finer, Nature Med. 2:714-6, 1996). The production of recombinant polypeptides in cultured mammalian cells is disclosed, for example, by Levinson et al., U.S. Patent No. 4,713,339; Hagen et al., U.S. Patent No. 4,784,950; Palmiter et al., U.S. Patent No. 4,579,821; and Ringold, U.S. Patent No. 4,656,134. Suitable cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK (ATCC No. CRL 1632), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., J. Gen. Virol. 36:59-72, 1977), NIH 3T3 fibroblasts (ATCC No. CRL-1658), Rat2 cells (ATCC No. CRL-1764), and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Manassas, VA. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Patent No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S. Patent Nos. 4,579,821 and 4,601,978) and the adenovirus major late promoter.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and are able to pass the gene of interest to their progeny are referred to as "stable transfectants." A preferred selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems can also be used to increase the expression

level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. A preferred amplifiable selectable marker is dihydrofolate reductase, which confers resistance to methotrexate. Other drug resistance genes (e.g. hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used. Alternative markers that introduce an altered phenotype, such as green fluorescent protein, or cell surface proteins such as CD4, CD8, Class I MHC, placental alkaline phosphatase may be used to sort transfected cells from untransfected cells by such means as FACS sorting or magnetic bead separation technology.

Other higher eukaryotic cells can also be used as hosts, including plant cells, insect cells and avian cells. The use of *Agrobacterium rhizogenes* as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., J. Biosci. (Bangalore) 11:47-58, 1987. Transformation of insect cells and production of foreign polypeptides therein is disclosed by Guarino et al., U.S. Patent No. 5,162,222 and WIPO publication WO 94/06463. Insect cells can be infected with recombinant baculovirus, commonly derived from *Autographa californica nuclear polyhedrosis virus* (AcNPV). See, King, L.A. and Possee, R.D., The Baculovirus Expression System: A Laboratory Guide, London, Chapman & Hall; O'Reilly, D.R. et al., Baculovirus Expression Vectors: A Laboratory Manual, New York, Oxford University Press., 1994; and, Richardson, C. D., Ed., Baculovirus Expression Protocols, Methods in Molecular Biology, Totowa, NJ, Humana Press, 1995. The second method of making recombinant baculovirus utilizes a transposon-based system described by Luckow (Luckow, V.A, et al., J Virol 67:4566-79, 1993). This system is sold in the Bac-to-Bac kit (Life Technologies, Rockville, MD). This system utilizes a transfer vector, pFastBac1™ (Life Technologies) containing a Tn7 transposon to move the DNA encoding the zmsc1 polypeptide into a baculovirus genome maintained in *E. coli* as a large plasmid called a "bacmid." The pFastBac1™ transfer vector utilizes the AcNPV polyhedrin promoter to drive the expression of the gene of interest, in this case zmsc1. However, pFastBac1™ can be modified to a considerable degree. The polyhedrin promoter can be removed and

substituted with the baculovirus basic protein promoter (also known as *Pcor*, p6.9 or MP promoter) which is expressed earlier in the baculovirus infection, and has been shown to be advantageous for expressing secreted proteins. See, Hill-Perkins, M.S. and Possee, R.D., J. Gen. Virol. 71:971-6, 1990; Bonning, B.C. et al., J. Gen. Virol. 75:1551-6, 1994; and, Chazenbalk, G.D., and Rapoport, B., J. Biol. Chem. 270:1543-9, 1995. In such transfer vector constructs, a short or long version of the basic protein promoter can be used. Moreover, transfer vectors can be constructed which replace the native zmse1 secretory signal sequences with secretory signal sequences derived from insect proteins. For example, a secretory signal sequence from Ecdysteroid Glucosyltransferase (EGT), honey bee Melittin (Invitrogen, Carlsbad, CA), or baculovirus gp67 (PharMingen, San Diego, CA) can be used in constructs to replace the native zmse1 secretory signal sequence. In addition, transfer vectors can include an in-frame fusion with DNA encoding an epitope tag at the C- or N-terminus of the expressed zmse1 polypeptide, for example, a Glu-Glu epitope tag (Grussenmeyer, T. et al., Proc. Natl. Acad. Sci. 82:7952-4, 1985). Using a technique known in the art, a transfer vector containing zmse1 is transformed into *E. coli*, and screened for bacmids which contain an interrupted lacZ gene indicative of recombinant baculovirus. The bacmid DNA containing the recombinant baculovirus genome is isolated, using common techniques, and used to transfect *Spodoptera frugiperda* cells, e.g. Sf9 cells. Recombinant virus that expresses zmse1 is subsequently produced. Recombinant viral stocks are made by methods commonly used the art.

The recombinant virus is used to infect host cells, typically a cell line derived from the fall armyworm, *Spodoptera frugiperda*. See, in general, Glick and Pasternak, Molecular Biotechnology: Principles and Applications of Recombinant DNA, ASM Press, Washington, D.C., 1994. Another suitable cell line is the High FiveO™ cell line (Invitrogen) derived from *Trichoplusia ni* (U.S. Patent No.5,300,435). Commercially available serum-free media are used to grow and maintain the cells. Suitable media are Sf900 II™ (Life Technologies) or ESF 921™ (Expression Systems) for the Sf9 cells; and Ex-cell0405™ (JRH Biosciences, Lenexa, KS) or Express FiveO™ (Life Technologies) for the *T. ni* cells. The cells are grown up from an inoculation density of approximately 2-5 x 10^5 cells to a density of 1-2 x 10^6 cells at

which time a recombinant viral stock is added at a multiplicity of infection (MOI) of 0.1 to 10, more typically near 3. Procedures used are generally described in available laboratory manuals (King, L. A. and Possee, R.D., ibid.; O'Reilly, D.R. et al., ibid.; Richardson, C. D., ibid.). Subsequent purification of the zmsel polypeptide from the supernatant can be achieved using methods described herein.

Fungal cells, including yeast cells, can also be used within the present invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia methanolica*. Methods for transforming *S. cerevisiae* cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for example, Kawasaki, U.S. Patent No. 4,599,311; Kawasaki et al., U.S. Patent No. 4,931,373; Brake, U.S. Patent No. 4,870,008; Welch et al., U.S. Patent No. 5,037,743; and Murray et al., U.S. Patent No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (e.g., leucine). A preferred vector system for use in *Saccharomyces cerevisiae* is the *POT1* vector system disclosed by Kawasaki et al. (U.S. Patent No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, e.g., Kawasaki, U.S. Patent No. 4,599,311; Kingsman et al., U.S. Patent No. 4,615,974; and Bitter, U.S. Patent No. 4,977,092) and alcohol dehydrogenase genes. See also U.S. Patents Nos. 4,990,446; 5,063,154; 5,139,936 and 4,661,454. Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia methanolica*, *Pichia guillermondii* and *Candida maltosa* are known in the art. See, for example, Gleeson et al., J. Gen. Microbiol. 132:3459-65, 1986 and Cregg, U.S. Patent No. 4,882,279. *Aspergillus* cells may be utilized according to the methods of McKnight et al., U.S. Patent No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino et al., U.S. Patent No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Patent No. 4,486,533.

The use of *Pichia methanolica* as host for the production of recombinant proteins is disclosed in WIPO Publications WO 97/17450, WO 97/17451, WO 98/02536, and WO 98/02565. DNA molecules for use in transforming *P. methanolica* will commonly be prepared as double-stranded, circular plasmids, which are preferably linearized prior to transformation. For polypeptide production in *P. methanolica*, it is preferred that the promoter and terminator in the plasmid be that of a *P. methanolica* gene, such as a *P. methanolica* alcohol utilization gene (*AUG1* or *AUG2*). Other useful promoters include those of the dihydroxyacetone synthase (DHAS), formate dehydrogenase (FMD), and catalase (CAT) genes. To facilitate integration of the DNA into the host chromosome, it is preferred to have the entire expression segment of the plasmid flanked at both ends by host DNA sequences. A preferred selectable marker for use in *Pichia methanolica* is a *P. methanolica ADE2* gene, which encodes phosphoribosyl-5-aminoimidazole carboxylase (AIRC; EC 4.1.1.21), which allows *ade2* host cells to grow in the absence of adenine. For large-scale, industrial processes where it is desirable to minimize the use of methanol, it is preferred to use host cells in which both methanol utilization genes (*AUG1* and *AUG2*) are deleted. For production of secreted proteins, host cells deficient in vacuolar protease genes (*PEP4* and *PRB1*) are preferred. Electroporation is used to facilitate the introduction of a plasmid containing DNA encoding a polypeptide of interest into *P. methanolica* cells. It is preferred to transform *P. methanolica* cells by electroporation using an exponentially decaying, pulsed electric field having a field strength of from 2.5 to 4.5 kV/cm, preferably about 3.75 kV/cm, and a time constant (t) of from 1 to 40 milliseconds, most preferably about 20 milliseconds.

Prokaryotic host cells, including strains of the bacteria *Escherichia coli*, *Bacillus* and other genera are also useful host cells within the present invention. Techniques for transforming these hosts and expressing foreign DNA sequences cloned therein are well known in the art (see, e.g., Sambrook et al., ibid.). When expressing a zmsc1 polypeptide in bacteria such as *E. coli*, the polypeptide may be retained in the cytoplasm, typically as insoluble granules, or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed, and the granules are recovered and denatured using, for example, guanidine isothiocyanate or

urea. The denatured polypeptide can then be refolded and dimerized by diluting the denaturant, such as by dialysis against a solution of urea and a combination of reduced and oxidized glutathione, followed by dialysis against a buffered saline solution. In the latter case, the polypeptide can be recovered from the periplasmic space in a soluble and functional form by disrupting the cells (by, for example, sonication or osmotic shock) to release the contents of the periplasmic space and recovering the protein, thereby obviating the need for denaturation and refolding.

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell. *P. methanolica* cells are cultured in a medium comprising adequate sources of carbon, nitrogen and trace nutrients at a temperature of about 25°C to 35°C. Liquid cultures are provided with sufficient aeration by conventional means, such as shaking of small flasks or sparging of fermentors. A preferred culture medium for *P. methanolica* is YEPD (2% D-glucose, 2% Bacto™ Peptone (Difco Laboratories, Detroit, MI), 1% Bacto™ yeast extract (Difco Laboratories), 0.004% adenine and 0.006% L-leucine).

It is preferred to purify the polypeptides of the present invention to ≥80% purity, more preferably to ≥90% purity, even more preferably ≥95% purity, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. Preferably, a purified polypeptide is substantially free of other polypeptides, particularly other polypeptides of animal origin.

Expressed recombinant zmsel polypeptides (or chimeric zmsel polypeptides) can be purified using fractionation and/or conventional purification

methods and media. For example, the zmsel polypeptides of the present invention can be purified using glutathione affinity chromatography followed by isopropyl-1-thio-β-D-galactopyranoside, such as that applied to other CRIB proteins (Burbelo, P.D. et al., J. Biol. Chem. 270:29071-29074, 1995). Moreover, other conventional purification methods can be used. Ammonium sulfate precipitation and acid or chaotrope extraction may be used for fractionation of samples. Exemplary purification steps may include hydroxyapatite, size exclusion, FPLC and reverse-phase high performance liquid chromatography. Suitable chromatographic media include derivatized dextrans, agarose, cellulose, polyacrylamide, specialty silicas, and the like. PEI, DEAE, QAE and Q derivatives are preferred. Exemplary chromatographic media include those media derivatized with phenyl, butyl, or octyl groups, such as Phenyl-Sepharose FF (Pharmacia), Toyopearl butyl 650 (Toso Haas, Montgomeryville, PA), Octyl-Sepharose (Pharmacia) and the like; or polyacrylic resins, such as Amberchrom CG 71 (Toso Haas) and the like. Suitable solid supports include glass beads, silica-based resins, cellulosic resins, agarose beads, cross-linked agarose beads, polystyrene beads, cross-linked polyacrylamide resins and the like that are insoluble under the conditions in which they are to be used. These supports may be modified with reactive groups that allow attachment of proteins by amino groups, carboxyl groups, sulfhydryl groups, hydroxyl groups and/or carbohydrate moieties. Examples of coupling chemistries include cyanogen bromide activation, N-hydroxysuccinimide activation, epoxide activation, sulfhydryl activation, hydrazide activation, and carboxyl and amino derivatives for carbodiimide coupling chemistries. These and other solid media are well known and widely used in the art, and are available from commercial suppliers. Methods for binding receptor polypeptides, or anti-complementary polypeptides, to support media are well known in the art. Selection of a particular method is a matter of routine design and is determined in part by the properties of the chosen support. See, for example, Affinity Chromatography: Principles & Methods, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1988.

The polypeptides of the present invention can be isolated by exploitation of their structural or biochemical properties. For example, immobilized metal ion adsorption (IMAC) chromatography can be used to purify histidine-rich proteins,

including those comprising polyhistidine tags. Briefly, a gel is first charged with divalent metal ions to form a chelate (Sulkowski, Trends in Biochem. 3:1-7, 1985). Histidine-rich proteins will be adsorbed to this matrix with differing affinities, depending upon the metal ion used, and will be eluted by competitive elution, lowering the pH, or use of strong chelating agents. Other methods of purification include purification of glycosylated proteins by lectin affinity chromatography and ion exchange chromatography (Methods in Enzymol., Vol. 182, "Guide to Protein Purification", M. Deutscher, (ed.), Acad. Press, San Diego, 1990, pp.529-39). Within additional embodiments of the invention, a fusion of the polypeptide of interest and an affinity tag (e.g., maltose-binding protein, an immunoglobulin domain) may be constructed to facilitate purification. Moreover purification methods used to purify mammalian, or other eukaryotic CRIB polypeptides can be used to purify human zmse1 polypeptides. See, for example, Burbelo, P.D. et al., supra.

Moreover, using methods described in the art, polypeptide fusions, or hybrid zmse1 proteins, are constructed using regions or domains of the inventive zmse1 in combination with those of other Rho effector family proteins (e.g. MSE55, PAK, WASP, other CRIB proteins, and the like), or heterologous proteins (Sambrook et al., ibid.; Altschul et al., ibid.; Picard, Cur. Opin. Biology, 5:511-5, 1994, and references therein). These methods allow the determination of the biological importance of larger domains or regions in a polypeptide of interest. Such hybrids may alter reaction kinetics, binding, constrict or expand the substrate specificity, alter activity in cytoskeletal reorganization or gene transcription in a cell, alter cytoskeletal organization and cell motility, transformation, or invasiveness, or alter tissue and cellular localization of a polypeptide, and can be applied to polypeptides of unknown structure.

Fusion proteins can be prepared by methods known to those skilled in the art by preparing each component of the fusion protein and chemically conjugating them. Alternatively, a polynucleotide encoding various components of the fusion protein in the proper reading frame can be generated using known techniques and expressed by the methods described herein. For example, part or all of a domain(s) conferring a structural or biological function may be swapped between zmse1 of the present invention with the functionally equivalent domain(s) from another family

member. Such domains include, but are not limited to, the CRIB motif, the N-terminal domain, C-terminal domain, or C-terminal tail. Such fusion proteins would be expected to have a biological functional profile that is the same or similar to polypeptides of the present invention or other known Rho effector family proteins (e.g. binding Cdc42, GTP hydrolysis or binding, increasing or decreasing actin polymerization, cell motility, or transformation, and the like) depending on the fusion constructed. Moreover, such fusion proteins may exhibit other properties as disclosed herein.

Standard molecular biological and cloning techniques can be used to swap the equivalent domains between the zmse1 polypeptide and those polypeptides to which they are fused. Generally, a DNA segment that encodes a domain of interest, e.g., a zmse1 active polypeptide or motif described herein, is operably linked in frame to at least one other DNA segment encoding an additional polypeptide and inserted into an appropriate expression vector, as described herein. Generally DNA constructs are made such that the several DNA segments that encode the corresponding regions of a polypeptide are operably linked in frame to make a single construct that encodes the entire fusion protein, or a functional portion thereof. For example, a DNA construct would encode from N-terminus to C-terminus a fusion protein comprising a signal polypeptide followed by a full length or mature polypeptide; or a DNA construct would encode from N-terminus to C-terminus a fusion protein comprising an N-terminal domain containing a CRIB motif and a C-terminal domain; or a DNA construct would encode from N-terminus to C-terminus a fusion protein comprising an N-terminal domain containing a CRIB motif; or, for example, any of the above as interchanged with equivalent regions from another protein. Such fusion proteins can be expressed, isolated, and assayed for activity as described herein.

Zmse1 polypeptides or fragments thereof may also be prepared through chemical synthesis. Zmse1 polypeptides may be monomers or multimers; glycosylated or non-glycosylated; pegylated or non-pegylated; and may or may not include an initial methionine amino acid residue.

Polypeptides of the present invention can also be synthesized by exclusive solid phase synthesis, partial solid phase methods, fragment condensation or

classical solution synthesis. Methods for synthesizing polypeptides are well known in the art. See, for example, Merrifield, J. Am. Chem. Soc. 85:2149, 1963; Kaiser et al., Anal. Biochem. 34:595, 1970. After the entire synthesis of the desired peptide on a solid support, the peptide-resin is with a reagent which cleaves the polypeptide from the resin and removes most of the side-chain protecting groups. Such methods are well established in the art.

The activity of molecules of the present invention can be measured using a variety of assays that measure actin polymerization, GTP binding or hydrolysis, proliferation, cell motility and invasion, metastasis or other CRIB/Rho effector protein activity. Of particular interest are assays that measure changes in cell proliferation, transformation, adhesion, gene expression, apoptosis, generation of nucleotide monophosphates, cell motility and invasion, metastasis, actin stress fiber production, filopodia, lamellipodia, membrane ruffling and others. Moreover CRIB/Rho effector protein activity can be measured using protein or antibody binding assays, scintillation proximity assay (SPA) technology described herein; cAMP assays described herein; as well as other assays described herein. Such assays are well known in the art, and many are described in further detail below.

As a CRIB protein zmse1 can affect cytoskeletal reorganization, cell-cell interaction and motility and hence can affect tissues that contract. Moreover, zmse1 is expressed in contractile tissues. For example contractile tissues in which zmse1 is expressed include testis, prostate, heart and skeletal muscle. The effects of zmse1 polypeptide, its antagonists and agonists, on tissue contractility can be measured *in vitro* using a tensiometer with or without electrical field stimulation. Such assays are known in the art and can be applied to tissue samples, such as aortic rings, muscle tissue, and other contractile tissue samples, as well as to organ systems, such as atria, and can be used to determine whether zmse1 polypeptide, its agonists or antagonists, enhance or depress contractility. Molecules of the present invention are hence useful for treating dysfunction associated with contractile tissues or can be used to suppress or enhance contractility *in vivo*. As such, molecules of the present invention have utility in treating cardiovascular disease, muscle relaxants or stimulants, infertility, *in vitro*

fertilization, birth control, treating impotence or other male reproductive dysfunction, as well as inducing birth.

The effect of the zmsel polypeptides, antagonists and agonists of the present invention on contractility of tissues including skeletal and smooth muscle tissues, testis, heart, and the like, can be measured in a tensiometer that measures contractility and relaxation in tissues. See, Dainty et al., J. Pharmacol. 100:767, 1990; Rhee et al., Neurotox. 16: 179, 1995; Anderson, M.B., Endocrinol, 114:364-368, 1984; and Downing, S.J. and Sherwood, O.D, Endocrinol. 116:1206-1214, 1985. For example, measuring vasodilatation of aortic rings is well known in the art. Briefly, aortic rings are taken from 4 month old Sprague Dawley rats and placed in a buffer solution, such as modified Krebs solution (118.5 mM NaCl, 4.6 mM KCl, 1.2 mM $MgSO_4.7H_2O$, 1.2 mM KH_2PO_4, 2.5 mM $CaCl_2.2H_2O$, 24.8 mM $NaHCO_3$ and 10 mM glucose). One of skill in the art would recognize that this method can be used with other animals, such as rabbits, other rat strains, Guinea pigs, and the like. The rings are then attached to an isometric force transducer (Radnoti Inc., Monrovia, CA) and the data recorded with a Powerlab physiology platform (Gould Instrument systems, Inc., Valley View, OH) and placed in an oxygenated (95% O_2, 5% CO_2) tissue bath containing the buffer solution. The tissues are adjusted to 1 gram resting tension and allowed to stabilize for about one hour before testing. The integrity of the rings can be tested with norepinepherin (Sigma Co., St. Louis, MO) and Carbachol, a muscarinic acetylcholine agonist (Sigma Co.). After integrity is checked, the rings are washed three times with fresh buffer and allowed to rest for about one hour. To test a sample for vasodilatation, or relaxation of the aortic ring tissue, the rings are contracted to two grams tension and allowed to stabilize for fifteen minutes. A zmsel polypeptide, antagonist or agonist sample is then added to 1, 2 or 3 of the 4 baths, without flushing, and tension on the rings recorded and compared to the control rings containing buffer only. Enhancement or relaxation of contractility by zmsel polypeptides, their agonists and antagonists is directly measured by this method, and it can be applied to other contractile tissues such as skeletal and smooth muscle tissue, gastrointestinal tissues, uterus, prostate, and testis.

The activity of molecules of the present invention can be measured using a variety of assays that measure stimulation of gastrointestinal cell contractility, modulation of nutrient uptake and/or secretion of digestive enzymes. Of particular interest are changes in contractility of smooth muscle cells, for example, the contractile response of segments of mammalian duodenum or other gastrointestinal smooth muscles tissue (Depoortere et al., J. Gastrointestinal Motility 1:150-159, 1989, incorporated herein by reference). An exemplary *in vivo* assay uses an ultrasonic micrometer to measure the dimensional changes radially between commissures and longitudinally to the plane of the valve base (Hansen et al., Society of Thoracic Surgeons 60:S384-390, 1995).

Gastric motility is generally measured in the clinical setting as the time required for gastric emptying and subsequent transit time through the gastrointestinal tract. Gastric emptying scans are well known to those skilled in the art, and briefly, comprise use of an oral contrast agent, such as barium, or a radiolabeled meal. Solids and liquids can be measured independently. A test food or liquid is radiolabeled with an isotope (e.g. ^{99m}Tc), and after ingestion or administration, transit time through the gastrointestinal tract and gastric emptying are measured by visualization using gamma cameras (Meyer et al., Am. J. Dig. Dis. 21:296, 1976; Collins et al., Gut 24:1117, 1983; Maughan et al., Diabet. Med. 13 9 Supp. 5:S6-10, 1996 and Horowitz et al., Arch. Intern. Med. 145:1467-1472, 1985). These studies may be performed before and after the administration of a promotility agent to quantify the efficacy of the drug.

An *in vivo* approach for assaying proteins of the present invention involves viral delivery systems. Exemplary viruses for this purpose include adenovirus, herpesvirus, retroviruses, vaccinia virus, and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acid (for review, see T.C. Becker et al., Meth. Cell Biol. 43:161-89, 1994; and J.T. Douglas and D.T. Curiel, Science & Medicine 4:44-53, 1997). The adenovirus system offers several advantages: (i) adenovirus can accommodate relatively large DNA inserts; (ii) can be grown to high-titer; (iii) infect a broad range of mammalian cell types; and (iv) can be used with many different promoters including ubiquitous, tissue specific, and regulatable promoters.

Also, because adenoviruses are stable in the bloodstream, they can be administered by intravenous injection.

Using adenovirus vectors where portions of the adenovirus genome are deleted, inserts are incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene has been deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell (the human 293 cell line is exemplary). When intravenously administered to intact animals, adenovirus primarily targets the liver. If the adenoviral delivery system has an E1 gene deletion, the virus cannot replicate in the host cells. However, the host's tissue (e.g., liver) will express and process (and, if a secretory signal sequence is present, secrete) the heterologous protein. Secreted proteins will enter the circulation in the highly vascularized liver, and effects on the infected animal can be determined.

Moreover, adenoviral vectors containing various deletions of viral genes can be used in an attempt to reduce or eliminate immune responses to the vector. Such adenoviruses are E1 deleted, and in addition contain deletions of E2A or E4 (Lusky, M. et al., J. Virol. 72:2022-2032, 1998; Raper, S.E. et al., Human Gene Therapy 9:671-679, 1998). In addition, deletion of E2b is reported to reduce immune responses (Amalfitano, A. et al., J. Virol. 72:926-933, 1998). Moreover, by deleting the entire adenovirus genome, very large inserts of heterologous DNA can be accommodated. Generation of so called "gutless" adenoviruses where all viral genes are deleted are particularly advantageous for insertion of large inserts of heterologous DNA. For review, see Yeh, P. and Perricaudet, M., FASEB J. 11:615-623, 1997.

The adenovirus system can also be used for protein production *in vitro*. By culturing adenovirus-infected non-293 cells under conditions where the cells are not rapidly dividing, the cells can produce proteins for extended periods of time. For instance, BHK cells are grown to confluence in cell factories, then exposed to the adenoviral vector encoding the secreted protein of interest. The cells are then grown under serum-free conditions, which allows infected cells to survive for several weeks without significant cell division. Alternatively, adenovirus vector infected 293 cells can be grown as adherent cells or in suspension culture at relatively high cell density to

produce significant amounts of protein (See Garnier et al., Cytotechnol. 15:145-55, 1994). With either protocol, an expressed, secreted heterologous protein can be repeatedly isolated from the cell culture supernatant, lysate, or membrane fractions depending on the disposition of the expressed protein in the cell. Within the infected 293 cell production protocol, non-secreted proteins may also be effectively obtained.

Moreover, the activity and effect of zmse1 on tumor progression and metastasis can be measured *in vivo*. Several syngeneic mouse models have been developed to study the influence of polypeptides, compounds or other treatments on tumor progression. In these models, tumor cells passaged in culture are implanted into mice of the same strain as the tumor donor. The cells will develop into tumors having similar characteristics in the recipient mice, and metastasis will also occur in some of the models. Appropriate tumor models for our studies include the Lewis lung carcinoma (ATCC No. CRL 1642) and B16 melanoma (ATCC No. CRL-6323), amongst others. These are both commonly used tumor lines, syngeneic to the C57BL6 mouse, that are readily cultured and manipulated *in vitro*. Tumors resulting from implantation of either of these cell lines are capable of metastasis to the lung in C57BL6 mice. The Lewis lung carcinoma model has recently been used in mice to identify an inhibitor of angiogenesis (O'Reilly MS, et al. Cell 79: 315-328,1994). C57BL*6J* mice are treated with an experimental agent either through daily injection of recombinant protein, agonist or antagonist or a one time injection of recombinant adenovirus. Three days following this treatment, 10^5 to 10^6 cells are implanted under the dorsal skin. Alternatively, the cells themselves may be infected with recombinant adenovirus, such as one expressing zmse1, before implantation so that the protein is synthesized at the tumor site or intracellularly, rather than systemically. The mice normally develop visible tumors within 5 days. The tumors are allowed to grow for a period of up to 3 weeks, during which time they may reach a size of 1500 - 1800 mm^3 in the control treated group. Tumor size and body weight are carefully monitored throughout the experiment. At the time of sacrifice, the tumor is removed and weighed along with the lungs and the liver. The lung weight has been shown to correlate well with metastatic tumor burden. As an additional measure, lung surface metastases are counted. The resected tumor, lungs and liver are prepared for histopathological

examination, immunohistochemistry, and *in situ* hybridization, using methods known in the art and described herein. The influence of the expressed polypeptide in question, e.g., zmse1, on the ability of the tumor to recruit vasculature and undergo metastasis can thus be assessed. In addition, aside from using adenovirus, the implanted cells can be transiently transfected with zmse1. Use of stable zmse1 transfectants as well as use of inducable promoters to activate zmse1 expression *in vivo* are known in the art and can be used in this system to assess zmse1 induction of metastasis. For general reference see, O'Reilly MS, et al. Cell 79:315-328, 1994; and Rusciano D, et al. Murine Models of Liver Metastasis. Invasion Metastasis 14:349-361, 1995.

The activation of zmse1 polypeptide can be measured by a silicon-based biosensor microphysiometer which measures the extracellular acidification rate or proton excretion associated with zmse1 activation and subsequent physiologic cellular responses. An exemplary device is the Cytosensor™ Microphysiometer manufactured by Molecular Devices, Sunnyvale, CA. A variety of cellular responses, such as cell proliferation, signal transduction, ion transport, energy production, inflammatory response, regulatory and receptor activation, and the like, can be measured by this method. See, for example, McConnell, H.M. et al., Science 257:1906-1912, 1992; Pitchford, S. et al., Meth. Enzymol. 228:84-108, 1997; Arimilli, S. et al., J. Immunol. Meth. 212:49-59, 1998; Van Liefde, I. Et al., Eur. J. Pharmacol. 346:87-95, 1998. The microphysiometer can be used for assaying adherent or non-adherent eukaryotic or prokaryotic cells. By measuring extracellular acidification changes in cell media over time, the microphysiometer directly measures cellular responses to various stimuli, including agonists, ligands, or antagonists of the zmse1 polypeptide. Preferably, the microphysiometer is used to measure responses of a zmse1-expressing eukaryotic cell, compared to a control eukaryotic cell that does not express zmse1 polypeptide. Zmse1-expressing eukaryotic cells comprise cells into which zmse1 has been transfected, as described herein, creating a cell that is responsive to zmse1-modulating stimuli; or cells naturally expressing zmse1. Differences, measured by a change in extracellular acidification, for example, an increase or diminution in the response of cells expressing zmse1, relative to a control, are a direct measurement of zmse1-modulated cellular responses. Moreover, such zmse1-modulated responses can be assayed under a variety

of stimuli. Also, using the microphysiometer, there is provided a method of identifying agonists and antagonists of zmse1 polypeptide, comprising providing cells expressing a zmse1 polypeptide, culturing a first portion of the cells in the absence of a test compound, culturing a second portion of the cells in the presence of a test compound, and detecting a change, for example, an increase or diminution, in a cellular response of the second portion of the cells as compared to the first portion of the cells. The change in cellular response is shown as a measurable change extracellular acidification rate. Antagonists and agonists, including the natural effectors or binding partners of zmse1 polypeptide, can be rapidly identified using this method.

In view of the protein family of which zmse1 is a member, agonists (including the natural ligand/ substrate/ cofactor/ etc.) and antagonists have enormous potential in both *in vitro* and *in vivo* applications. Compounds identified as zmse1 agonists and antagonists are useful for modulating tumor cell motility, invasion, and metastasis, modulating actin polymerization and cytoskeletal reorganization, gene transcription, modulating contractility of various tissues as described herein, modulating proliferation (e.g., of cancerous cells), modulating digestion, modulating heart conditions, modulating testicular function and fertility, and the like *in vitro* and *in vivo*. For example, zmse1 and agonist or antagonist compounds are useful as components of defined cell culture media, and may be used alone or in combination with other cytokines and hormones to replace serum that is commonly used in cell culture. Agonists or antagonists are thus useful in specifically promoting the growth and/or development of cell lineages in culture. Alternatively, zmse1 polypeptides and zmse1 agonist or antagonist polypeptides are useful as a research reagent, such as for the expansion of cell lines, or useful as an amino acid source for cell culture.

The activity of molecules of the present invention can be measured using a variety of assays that measure proliferation and/or differentiation of specific cell types, chemotaxis, adhesion, changes in ion channel influx, pH flux, regulation of second messenger levels and neurotransmitter release, cell motility, protein binding, apoptosis, or the like. Such assays are well known in the art. See, for example, in "Basic & Clinical Endocrinology Ser., Vol. 3," Cytochemical Bioassays: Techniques & Applications, Chayen; Chayen, Bitensky, eds., Dekker, New York, 1983.

The activity of molecules of the present invention can also be measured using a variety of assays that measure, for example, signal transduction upon binding a ligand or substrate, or antibody binding to the outside of an intact cell and stimulating the signal transduction pathway of zmse1. For example, zmse1 polypeptides, complementary binding polypeptides, or anti-zmse1 antibodies can be labeled and tested for specific and saturating binding to specific substrates, cell lines or cells. Identification of positive cells to which zmse1 polypeptides, complementary binding polypeptides, or anti-zmse1 antibodies binds can be achieved by injecting a radioactively or fluorescently-labeled zmse1 polypeptide, polypeptide fragments, complementary binding polypeptides, or anti-zmse1 antibodies and using art-recognized immunohistochemistry methods to visualize a cell or tissue *in vivo* where zmse1 binds or is expressed. After identification of bound positive cells, activity can be tested for zmse1-mediated activation of a signal transduction pathway using methods known in the art. For instance, vector constructs containing a reporter (e.g. SRE-luciferase, STAT-luciferase, thyroid hormone response element (THRE)-luciferase, SV40 promoter-luciferase or the like) can be introduced into the positive cell lines expressing zmse1; such cell lines, when exposed to conditioned media containing secreted zmse1 activating proteins will demonstrate zmse1-mediated signal transduction activity through activation of the measurable reporter. Such assays are well known in the art. Specific assays include, but are not limited to, bioassays measuring signal transduction.

The activity of molecules of the present invention can also be measured using a variety of assays that measure, for example, cell motility, adhesion and invasion *in vitro* and metastasis *in vivo*. Such assays are known in the art. For example, motility assays in NIH 3T3 cells, mouse keratinocytes, and epithelial cells are described in Takiashi, K. et al., Mol. Cell Biol. 13:72-79, 1993; Takiashi, K. et al., Oncogene 8:273-278, 1994; Ridley, A.J. et al., Mol. Cell Biol. 15:1110-1122, 1995; and Keely, P.J. et al., Nature 360:632-636, 1997. For review and application of *in vitro* invasion assays; for example, using hepatoma or lymphoma cells invasion through mesothelial or fibroblast cell monolayers, phagokenesis and wound healing assays. For example, see Yoshioka, K. et al., FEBS Lett. 372:25-28, 1995; Wang, W.Z., and Ron D. Science

272:1347-1349, 1996; Habets, G. Cell 77:537-549, 1994; and Michiels, F. et al., Nature 375:338-340, 1995; Michiels, F. and Collard, J.G., Biochem. Soc. Symp. 65:215-146, 1999; and Keely, P.J. supra,. Moreover, *in vivo* metastasis assays can be used to assess zmse1 polypeptide, expression, agonist or antagonist activity *in vivo* in mice (Verschueren, H. Eur. J. Cell Biol. 73:182-187, 1997). Cell adhesion can be assessed by the adherence or non-adherence of normally adherent cell lines to cell culture dishes, amongst other assays known in the art.

Moreover, the activity of molecules of the present invention can also be measured using a variety of assays that measure cytoskeletal reorganization. Such assays are well known in the art. For example, effects of zmse1 on membrane ruffling can be assessed in Swiss 3T3 cells (Ridley, A.J. Cell 70:401-410, 1992). Actin polymerization and cytoskeletal rearrangement including assessment of actin stress fibers, focal complexes, lamellipodia and filopodia, can be assessed by various means including immunofluorescence, and time-lapse imaging amongst other known methods (Symons, M. et al., Cell 84:723-734, 1996; Nobes, C.D., and Hall, A., Cell 81:53-62, 1995; Burbelo, P.D. et al., Proc. Natl. Acad. Sci. USA 96:9083-9088, 1999; Aspenstrom, P. Exper. Cell. Res. 246:20-25, 1999; Gallo, G., and Letourneau, P.C., Current Biol. 8:R80-R82, 1998; and Miki, H. et al., Nature 391:93-96, 1998).

Moreover, the activity of molecules of the present invention can also be measured using a variety of assays that measure protein binding. For example, the zmse1 polypeptides of the present invention can be assessed for their ability to bind Rho family proteins, such as Cdc42, Rac and Rho, in filter binding assays with GST-rhoGAP as a positive control (Burbelo, P.D. et al., J. Biol. Chem. 270:29071-29074, 1995; and Lancaster, C.A. et al., J. Biol. Chem. 269:1137-1142, 1994. Moreover, guanine nucleotide dependence of such binding can also be determined using a glutathione-agarose bead assay or other method known in the art (for example, see Burbelo, P.D. et al., supra,; and Burbelo, P.D. et al., Proc. Natl. Acad. Sci. USA 96:9083-9088, 1999). Moreover, GTP hydrolysis can be measured as a product of zmase1 activity. Such assays are known in the art.

Zmse1 can also be used to identify modulators (e.g. agonists or antagonists) of its activity. Test compounds are added to the assays disclosed herein to

identify compounds that inhibit or stimulate the activity of zmse1. In addition to those assays disclosed herein, samples can be tested for inhibition/stimulation of zmse1 activity within a variety of assays designed to measure zmse1 binding, dimerization, heterodimerization, DNA binding or the stimulation/inhibition of zmse1-dependent cellular responses. For example, zmse1-expressing cell lines can be transfected with a reporter gene construct that is responsive to a zmse1-stimulated cellular pathway. Reporter gene constructs of this type are known in the art, and will generally comprise a zmse1-DNA response element operably linked to a gene encoding an assay detectable protein, such as luciferase. DNA response elements can include, but are not limited to, cyclic AMP response elements (CRE), hormone response elements (HRE) insulin response element (IRE) (Nasrin et al., Proc. Natl. Acad. Sci. USA 87:5273-7, 1990) and serum response elements (SRE) (Shaw et al. Cell 56: 563-72, 1989). Cyclic AMP response elements are reviewed in Roestler et al., J. Biol. Chem. 263:9063-6; 1988 and Habener, Molec. Endocrinol. 4:1087-94; 1990. Hormone response elements are reviewed in Beato, Cell 56:335-44; 1989. Candidate compounds, solutions, mixtures or extracts or conditioned media from various cell types are tested for the ability to enhance the activity of zmse1 signal transduction as evidenced by an increase in zmse1 stimulation of reporter gene expression. Assays of this type will detect compounds that directly stimulate zmse1 signal transduction activity through binding the upstream receptor or by otherwise stimulating part of the signal cascade in which zmse1 is involved. As such, there is provided a method of identifying agonists of zmse1 polypeptide, comprising providing cells expressing zmse1 responsive to a zmse1 pathway, culturing a first portion of the cells in the absence of a test compound, culturing a second portion of the cells in the presence of a test compound, and detecting a increase in a cellular response of the second portion of the cells as compared to the first portion of the cells. Moreover a third cell, containing the reporter gene construct described above, but not expressing zmse1 polypeptide, can be used as a control cell to assess non-specific, or non-zmse1-mediated, stimulation of the reporter. Agonists are useful to stimulate or increase zmse1 polypeptide function.

Moreover, compounds or other samples can be tested for direct blocking of zmse1 binding to another protein or substrate, e.g., a heterodimer described below,

using zmse1 tagged with a detectable label (e.g., ^{125}I, biotin, horseradish peroxidase, FITC, and the like). Within assays of this type, the ability of a test sample to inhibit the binding of labeled zmse1 to the other protein or substrate is indicative of inhibitory activity, which can be confirmed through secondary assays. Proteins used within binding assays may be cellular proteins or isolated, immobilized proteins.

Zmse1 activation can be detected by: (1) measurement of adenylate cyclase activity (Salomon et al., Anal. Biochem. 58:541-48, 1974; Alvarez and Daniels, Anal. Biochem. 187:98-103, 1990); (2) measurement of change in intracellular cAMP levels using conventional radioimmunoassay methods (Steiner et al., J. Biol. Chem. 247:1106-13, 1972; Harper and Brooker, J. Cyc. Nucl. Res. 1:207-18, 1975); or (3) through use of a cAMP scintillation proximity assay (SPA) method (Amersham Corp., Arlington Heights, IL). These methods provide sensitivity and accuracy.

An alternative assay system involves selection of polypeptides that are able to induce expression of a cyclic AMP response element (CRE)-luciferase reporter gene, as a consequence of elevated cAMP levels, in cells expressing a zmse1 polypeptide, but not in cells lacking zmse1 expression, analogous to such assays employing calcitonin receptor as described in U.S. patent No. 5,622,839, U.S. Patent No. 5,674,689, and U.S. patent No. 5,674,981.

In addition, polypeptides of the present invention can be assayed and used for their ability to modify inflammation. As zmse1 may induce cell migration and/or affect contractility in tissues, it may be involved in migration of inflammatory cells. Methods to determine proinflammatory and anti-inflammatory qualities of zmse1 polypeptide, its agonists or antagonists, are known in the art and discussed herein. For example, suppression of cAMP production is an indication of anti-inflammatory effects (Nihei, Y., et al., Arch. Dermatol. Res., 287:546-552, 1995). Suppression of cAMP and inhibition of ICAM and HLA-Dr induced by IFN-γ in keratinocytes can be used to assess inhibition of inflammation. Alternatively, enhancement of cAMP production and induction of ICAM and HLA-Dr in this system can be an measurement of proinflammatory effects of a protein. As a member of a signal transduction cascade, zmse1, likewise can exhibit similar inflammatory effects, and may exert these effects in tissues in which it is expressed, or indirectly in other tissues. For example, zmse1 is